

Fig. 2

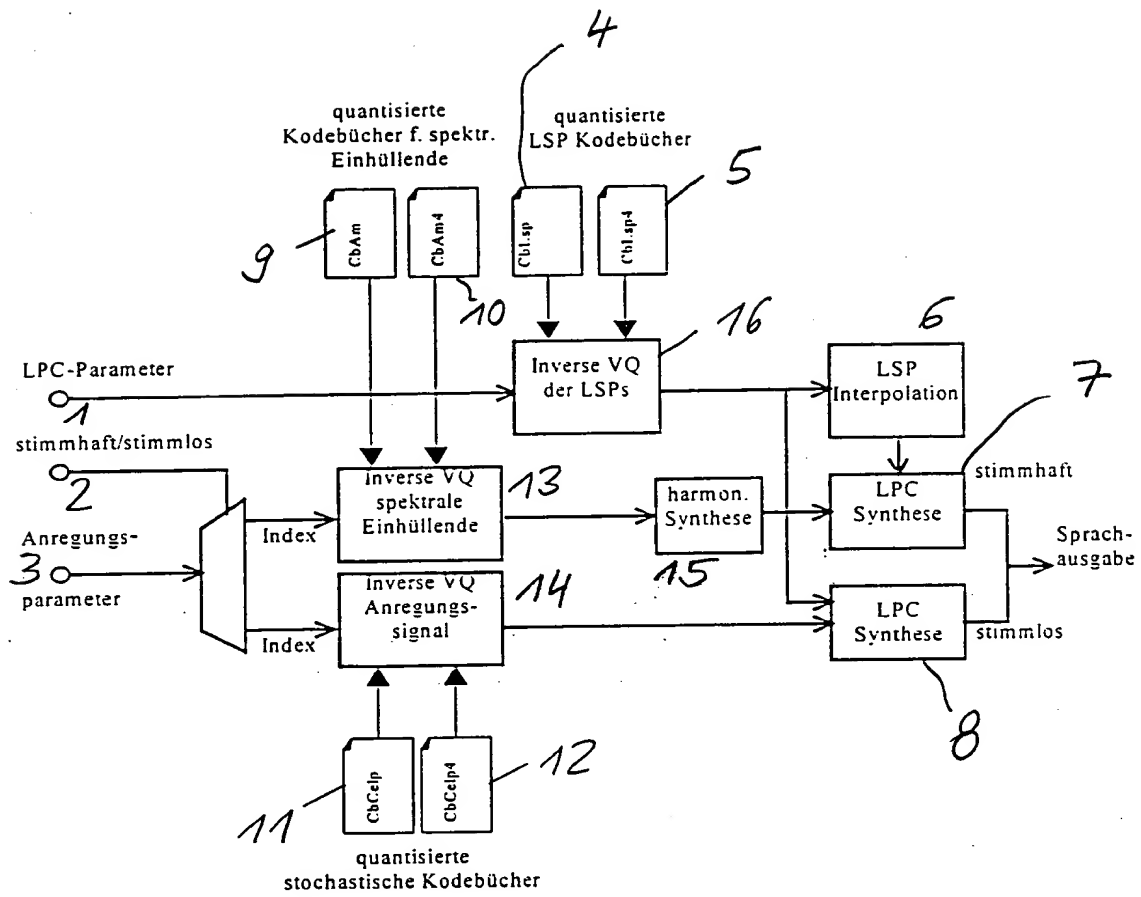


Fig. 1

06.10.98 Sk/Hy

5

ROBERT BOSCH GMBH, 70442 Stuttgart

Verfahren zur Codierung oder Decodierung von
Sprachsignalabtastwerten sowie Coder bzw. Decoder

10

Zusammenfassung

15

Zur Codierung oder Decodierung von Sprachsignalabtastwerten werden die in den Codebüchern/Codetabellen enthaltenen Werte zur Generierung der Sprachsignalparameter in quantisierter Form abgelegt.

20

Ohne Verschlechterung der Sprachqualität kann die Verarbeitung mit Prozessoren mit Ganzzahlverarbeitung durchgeführt werden.

Fig. 1

quantisierter Form in den entsprechenden Codebüchern/
Tabellen abgelegt werden.

5

11. Verfahren nach einem der Ansprüche 1 bis 7, dadurch
gekennzeichnet, daß für einen CELP (Code Excited Linear
Prediction) - Sprachcoder/ -decoder die Werte für die LSP
(Line Spectral Pairs) -VQ-Vektorquantisierungs-Codebuch-/
Tabelleneinträge sowie die der Gain-VQ Tabelleneinträge in
quantisierter Form abgelegt werden.

10

12. Coder oder Decoder für die Verarbeitung von
Sprachsignalabtastwerten unter Verwendung der Analyse durch
Synthese - Methode mit folgenden Maßnahmen:

15

die in den Codebüchern/Codetabellen (4,5,9,10,11,12,25,26)
enthaltenen Werte zur Generierung der Sprachsignalparameter
sind in quantisierter Form abgelegt, wobei die Wortbreite so
gewählt ist, daß keine merklichen Verluste der
Sprachqualität auftreten.

Faktor kleiner als Eins, vorzugsweise 0,5, vorgenommen wird und daß diese Multiplikation so oft wiederholt wird, bis alle Elemente im Wertebereich liegen.

5 5. Verfahren nach Anspruch 4, dadurch gekennzeichnet, daß die Anzahl der wiederholten Multiplikationen als Skalierfaktor für alle Codebuch-/ Tabelleneinträge verwendet wird.

10 6. Verfahren nach Anspruch 5, dadurch gekennzeichnet, daß eine Skalierung der Codebuch-/ Tabelleneinträge auf die Bits des erforderlichen Wertebereiches vorgenommen wird.

15 7. Verfahren nach Anspruch 6, dadurch gekennzeichnet, daß für eine endgültige Quantisierung eine Rundung und eine nachfolgende Abschneidung von Nachkommastellen erfolgt.

20 8. Verfahren nach einem der Ansprüche 1-7, dadurch gekennzeichnet, daß die Wortbreite der quantisierten Werte zu 16 Bit gewählt wird.

25 9. Verfahren nach einem der Ansprüche 1 bis 8, dadurch gekennzeichnet, daß die Verarbeitung der quantisierten Codebuch-/ Tabelleneinträge mittels digitaler Signalverarbeitung im Ganzzahlformat vorgenommen wird.

30 10. Verfahren nach einem der Ansprüche 1 bis 9, dadurch gekennzeichnet, daß für einen HVXC (Harmonic Vector Excitation Coding) - Sprachcoder/ Sprachdecoder die LPC-Koeffizienten, die spektralen Einhüllenden des Sprachsignal und der stimmlosen Abschnitte des Sprachsignals in

06.10.98 Sk

5

ROBERT BOSCH GMBH, 70442 Stuttgart

Ansprüche

10

1. Verfahren zur Codierung oder Decodierung von Sprachsignalabtastwerten insbesondere unter Verwendung der Analyse durch Synthese - Methode mit folgenden Schritten:

15

- die für die Generierung der Sprachsignalparameter verwendeten zuvor aus den Sprachsignalabtastwerten analysierten Werte werden vor ihrer Abspeicherung in Codebüchern/Codetabellen quantisiert,
- die Quantisierung der Werte erfolgt auf eine Wortbreite, die zu keinen merklichen Verlusten der Sprachqualität führt.

20

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die Wortbreiten der in den Codebüchern/Codetabellen abgelegten Werte durch Hörtests ermittelt werden.

25

3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß die Werte eines jeden Codebuches/ einer jeden Codetabelle so skaliert werden, daß der zur Verfügung stehende Wertebereich möglichst komplett ausgenutzt wird.

30

4. Verfahren nach Anspruch 3, dadurch gekennzeichnet, daß zur Skalierung das Maximum der positiven und negativen Werte eines jeden Codebuchs/Codetabelle ermittelt wird, daß im Falle des Überschreitens des zur Verfügung stehenden Wertebereiches eine Multiplikation der Werte mit einem

Unterrahmen die Interpolation - Baugruppe 28 - zwischen den LSP-Parametern des vergangenen und des aktuellen Rahmens. Die in LPC-Parameter umgewandelten LSP-Parameter gehen als Koeffizienten in das LPC-Synthesefilter 29 ein. Dort erfolgt
5 die Rekonstruktion der Sprachdaten durch Filterung des Erregungssignals. Zur Verbesserung der Sprachqualität kann das rekonstruierte Sprachsignal noch in einem Postfilter 30 gefiltert werden.

10 Die LSP-VQ-Tabellenwerte sowie die Gain-VQ-Tabellenwerte für die Codebücher 25 und 26, die zuvor aus den Sprachsignalabtastwerten analysiert wurden, liegen normalerweise in einer Gleitkommadarstellung vor, die wie
15 zuvor erläutert für eine Fixpunkt-DSP Verarbeitung nicht geeignet ist. Es erfolgt aus den gleichen Gründen wie beim HVXC-Decoder (Fig. 1) eine Umwandlung der Tabellenwerte in eine quantisierte Form. Die Verfahrensschritte bei dieser Quantisierung wie insbesondere die Ermittlung des Wertebereichs für die Codebücher erfolgt wie bei der zuvor
20 erläuterten Quantisierung.

Die bisherigen Ausführungsbeispiele der Erfindung wurden anhand von Sprachdecodern erläutert. Natürlich kann die Erfindung auch bei entsprechenden Codern (Encodern)
25 eingesetzt werden, die Codebücher verwenden. Auch dort können die Codebucheinträge für die Aufbereitung von Sprachsignalen für die Übertragung zuvor quantisiert werden. Beispiele von solchen Encodern, deren Codebucheinträge zuvor quantisiert werden können sind aus EP 0545 386 A2, US
30 5,208,862, US 5,487,128, US 5,199,076 oder US 5,261,027 bekannt.

Eine weitere Ausgestaltung der Erfindung wird in Zusammenhang mit Fig. 2 erläutert. Dort ist das Blockschaltbild eines CELP-Decoders dargestellt. Zunächst werden die zur Decodierung eines Rahmens notwendigen Elemente wie zuvor aus einem übertragenen Bitstrom gelesen. Dabei handelt es sich um die LPC Indizes, die Erregungsparameter (Lag und Shape Index) sowie die Amplituden Indizes (Gain Indices). Angeliefert werden diese Parameter (Elemente) an den Decodereingängen 17 bis 21. Die Erregungsparameter setzen sich aus den Parametern für das adaptive Codebuch (Lag) 22 zur Generierung periodischer Signalkomponenten (stimmhaft) und den Parametern für die festen Codebücher (Shape Index) 23a23n zusammen.

Die Einträge der festen Codebücher 23a ...23n und des adaptiven Codebuchs 22 werden jeweils mit einem Skalierungsfaktor (Gain) über den Gain-Decoder 24 multipliziert. Dieser Skalierungsfaktor wird unter Zuhilfenahme der Gain Indizes, die am Eingang 21 anliegen, und der Gain-VQ (Vektorquantisierung)-Tabellen, die in den Codebüchern 25 abgelegt sind, rekonstruiert. Der endgültige Erregungsvektor setzt sich aus der Summe der festen und des adaptiven Codebuchvektors zusammen.

Bei der Nutzung des Vektorquantisierers VQ repräsentieren die LPC-Indizes die vektorquantisierten LSP-Parameter (Line Spectral Pairs). Die Vektoren der ersten und zweiten Stufe der inversen Vektorquantisierung der LSP-Parameter erhält man durch Auslesen der LSP-VQ-Tabellenwerte, die in den Codebüchern 26 abgespeichert sind. Die endgültige Rekonstruktion der LPC-Parameter erfolgt im LPC-Parameter-Decoder 27. Innerhalb eines jeden Rahmens erfolgt für jeden

Nach diesem Schritt liegen die Einträge von jedem Kodebuch im folgenden Wertebereich: $-1 \leq b_n \leq (1 - 2^{-(wordlength-1)})$, mit $0 \leq n \leq m$.

5 3.) Skalierung auf wordlength Bits.

Zur Skalierung auf den erforderlichen Wertebereich, erfolgt die Multiplikation mit $2^{wordlength-1}$. Damit liegen die Werte der Kodebücher c_n im Bereich zwischen $-2^{wordlength-1}$ und $2^{wordlength-1} - 1$.

4.) Runden

Bevor die Nachkommastellen abgeschnitten werden, erfolgt die Rundung der ermittelten Einträge. Dazu wird in Abhängigkeit des Vorzeichens +0,5 bzw. -0,5 addiert. Dieses geschieht in der folgenden Form:

15 $c_n \geq 0 : d_n = c_n + 0,5$

$c_n < 0 : d_n = c_n - 0,5$

Hierbei ist zu beachten, daß der maximal zulässige Wertebereich nicht überschritten wird. Dieser liegt in dem Bereich, wie unter 2.) angegeben.

5.) Abtrennen der Nachkommastellen

Die endgültige Quantisierung erfolgt durch das Abtrennen der Nachkommastellen. Damit erhält man die quantisierten Werte.

25 Versuche haben gezeigt, daß mit der Festlegung der Variablen wordlength auf 16 eine vom Original nicht zu unterscheidende Sprachqualität erhalten wird.

$max_pos = \max(\{a_n \in cb | a_n \geq 0\})$ bzw. $max_neg = \min(\{a_n \in cb | a_n < 0\})$,
mit $0 \leq n \leq m$.

5 In Abhängigkeit der Größe von max_pos bzw. max_neg , ergeben sich die folgenden Schritte:

$max_pos > (1 - 2^{-(wordlength-1)})$ oder $max_neg < -1$

max_pos und max_neg werden mit $\frac{1}{2}$ multipliziert. Erfüllt das Resultat immer noch die unter (a) gestellte Bedingung, dann muß der Vorgang wiederholt werden, bis die Bedingung nicht mehr zutrifft. Die Anzahl der Multiplikationen mit $\frac{1}{2}$ wird gezählt und in der Variablen *scale* abgelegt.

$max_pos \leq (1 - 2^{-(wordlength-1)})$ oder $max_neg \geq -1$

15 max_pos und max_neg werden mit 2 multipliziert. Erfüllt das Resultat immer noch die unter (b) gestellte Bedingung, dann muß der Vorgang wiederholt werden, bis die Bedingung nicht mehr zutrifft. Die Anzahl der Multiplikationen mit 2 wird gezählt und in der Variablen *scale* abgelegt.

2.) Skalierung der Elemente von *cb* auf den Bereich zwischen -1 und $(1 - 2^{-(wordlength-1)})$.

In Abhängigkeit der unter 1.) getroffenen Entscheidung erfolgt die Skalierung aller Kodebucheinträge auf den genannten Bereich:

$$b_n = \frac{1}{2^{scale}} a_n \forall a_n \in cb \quad \text{mit } 0 \leq n \leq m$$

25 $b_n = 2^{scale} a_n \forall a_n \in cb \quad \text{mit } 0 \leq n \leq m.$

die einzelnen Tabellenwerte werden in verschiedenen Hörtests ermittelt.

5 Die Quantisierung erfolgt auf eine Wortbreite, die in verschiedenen Tests ermittelt wird. In der folgenden Darstellung wird diese Wortbreite allgemein mit *wordlength* bezeichnet. Diese Größe wird in Bits ausgedrückt. Eine vorzeichenbehaftete ganze Zahl mit *wordlength* Bits umfaßt einen Wertebereich von $-2^{\text{wordlength}-1}$ bis $2^{\text{wordlength}-1}-1$. Die Quantisierung der Kodebücher erfolgt damit auf die nachfolgend gezeigte Art. Den Ausgangspunkt stellen die in „study on ISO/IEC 14496-3 FCD, Subpart 3“, definierten Kodebücher dar. Das Kodebuch *cb* wird für dieses Dokument wie folgt definiert: $cb = \{a_0, a_1, \dots, a_n, \dots, a_m\}$ mit $0 \leq n \leq m$ und $a_n \in R$.
15 Zur Quantisierung der einzelnen Elemente sind die folgenden Schritte erforderlich:

1.) Ermittlung des Wertebereichs der Kodebücher

20 Um eine gut angepaßte Quantisierung zu erhalten, werden die Elemente eines jeden Kodebuchs so skaliert, daß der zur Verfügung stehende Wertebereich möglichst komplett ausgenutzt wird. Dazu muß der Wertebereich der Elemente zwischen

25 $\frac{-2^{\text{wordlength}-1}}{2^{\text{wordlength}-1}} = -1$ und $\frac{2^{\text{wordlength}-1}-1}{2^{\text{wordlength}-1}} = 1 - 2^{-(\text{wordlength}-1)}$

liegen. Um dies zu erreichen, wird das Maximum der positiven und der negativen Elemente (*max_pos* bzw. *max_neg*) eines jeden Kodebuchs ermittelt. Diese ergeben sich aus

gegebenenfalls die Interpolation - Baugruppe 6 - zwischen den LSP-Parametern des vergangenen und aktuellen Rahmens, womit eine Aktualisierung dieser Werte in einem Raster von 2,5 ms erreicht wird. Im Anschluß daran erfolgt die
5 Umwandlung in LPC-Parameter, die als Koeffizienten in das LPC-Synthesefilter - Baugruppen 7 und 8 - eingehen.

Parallel zu dieser Berechnung und in Abhängigkeit der stimmhaft/stimmlos-Entscheidung werden die Vektoren für die spektrale Einhüllende (stimmhafter Rahmen), AM-Codebücher 9 (CbAm) und 10 (CbAm4) bzw. die Vektoren für das stochastische Anregungssignal (stimmloser Rahmen, CELP-Codebücher 11 (CbCelp) und 12 (CbCelp4)) gelesen. Die Regenerierung der spektralen Einhüllenden und des
15 Anregungssignals erfolgt mit den inversen Vektorquantisierern 13 und 14. Nach der harmonischen Synthese (stimmhaft) - Baugruppe 15 - erfolgt die Filterung der Sprachdaten im LPC-Synthesefilter. Die Ausgangsdaten aus dem stimmhaften - Baugruppe 7 - und dem stimmlosen -
20 Baugruppe 8 - Synthesefilter werden abschließend addiert, womit das rekonstruierte Sprachsignal für einen Rahmen von 20 ms vorliegt.

Da sich, wie zuvor erläutert, Werte für die Codebücher in
25 Gleitkommadarstellung nicht für

Fixpunkt-DSPs eignen, weil die erforderlichen Wortbreiten zu groß wären (Speicherbedarf, interne Wortbreiten und Arithmetik, ROM), erfolgt die Umsetzung der Tabellenwerte für die Codebücher, die zuvor aus den
30 Sprachsignalabtastwerten analysiert wurden, in eine quantisierte Form bei resultierender äquivalenter Sprachqualität. Die hierfür erforderlichen Wortbreiten für

beträchtliche Einsparung von Speicherkapazität, insbesondere in Form von ROMs möglich. Die Erfindung ist bei unterschiedlichen Sprachsignalcodierverfahren einsetzbar, beispielsweise für HVXC-Coder/Decoder oder CELP-Coder/Decoder.

Zeichnungen

Anhand der Zeichnungen werden Ausführungsbeispiele der Erfindung näher erläutert. Es zeigen Fig.1 ein vereinfachtes Blockschaltbild eines HVXC-Sprachdecoders, Fig.2 ein vereinfachtes Blockschaltbild eines CELP-Sprachdecoders.

Beschreibung von Ausführungsbeispielen

Bevor auf die eigentliche Quantisierung eingegangen wird, wird zunächst ein Sprachdecoder vorgestellt, bei dem die erfindungsgemäße Quantisierung eingesetzt wird. Bei dem HVXC-Sprachdecoder nach Fig.1 werden die übertragenen Sprachparameter, nämlich die LPC-Parameter, die stimmhaft/stimmlos-Entscheidung des Encoders und die Anregungsparameter, die in einem Übertragungsrahmen von 20 ms Dauer untergebracht sind, aus dem Bitstrom gelesen und als Eingangssignale an den Eingängen 1, 2 und 3 angeliefert.

Die LPC-Parameter enthalten Indizes, aus denen der inverse LSP-Vektorquantisierer 16 die LSP-Parameter (Line Spectral Pairs) regeneriert. Dazu werden die LSP-Codebücher 4 (CbLsp) und 5 (CbLsp4) mit den LPC-Parametern indiziert und die LSP-Parameter ausgelesen. In Abhängigkeit der stimmhaft/stimmlos-Entscheidung dieses Rahmens erfolgt

Zur Codierung solcher Sprachsignale wird häufig die „Analyse durch Synthese“-Methode verwendet (ANT Nachrichtentechnische Berichte Heft 5, Nov. 1988, Seiten 93 bis 105). Bei den erwähnten Sprachcodierverfahren werden in Codebüchern, d.h. in den Tabellen, Werte abgespeichert, die für die Generierung der Signalparameter und damit für die Koeffizienten der Sprachsynthesefilter verwendet werden. Über eine Indexsteuerung werden die in den Codebüchern abgespeicherten Werte ausgelesen.

Vorteile der Erfindung

Durch die Maßnahmen des Anspruchs 1, d.h. insbesondere durch die Quantisierung der Werte in den Codebüchern, werden die vorliegenden Daten in ihrer Genauigkeit (Quantisierung) beschränkt, so daß die Codebucheinträge mit einer endlichen Wortbreite dargestellt werden können. Somit kann ihre Portierung auf digitale Signalprozessoren mit Ganzzahlarithmetik erfolgen, ohne die durch Standards, insbesondere gemäß ISO/IEC 14496-3, vorgegebenen Qualitätsanforderungen zu verletzen. Im Gegensatz zur Erfindung liegen in den erwähnten Arbeitsversionen der Standards die Werte für die Codebücher unquantisiert im Gleitkommaformat vor und können nur mit sehr aufwendigen und speicherintensiven Verfahren direkt verarbeitet werden.

Trotz der Genauigkeitsbeschränkung der Tabellenwerte ist bei der Erfindung eine gleiche subjektive Qualität nach der Sprachdecodierung zu erzielen. Mit den Maßnahmen der Erfindung ist eine leichte und standardkonforme Portierung des Codes auf unterschiedliche Rechnerplattformen möglich, ohne Beeinflussung der subjektiven Qualität des Coders. Da reduzierte Wortbreiten verwendet werden, ist eine

06.10.98 Sk/Hy

5

ROBERT BOSCH GMBH, 70442 Stuttgart

10

Verfahren zur Codierung oder Decodierung von
Sprachsignalabtastwerten sowie Coder bzw. Decoder

Stand der Technik

15

Die Erfindung geht aus von einem Verfahren zur Codierung
oder Decodierung von Sprachsignalabtastwerten.

20

Im Standard zur Codierung Audiovisueller Objekte nach MPEG-4
sind in ISO/IEC 14496-3 FCD, Subpart 2 parametrische Coder
beschrieben, insbesondere der HVXC-Coder (Harmonic Vector
Excitation Coding) zur Codierung von Sprache bei extrem
niedrigen Bitraten. Dieser Standard enthält zur Generierung
der LPC-Koeffizienten, der spektralen Einhüllenden des
Sprachsignals und der stimmlosen Abschnitte mehrere Tabellen,
die im Gleitkommaformat vorliegen.

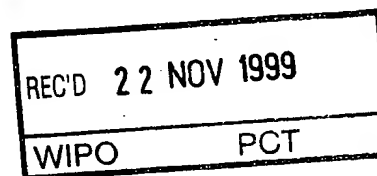
25

Im Subpart 3 dieses Standards wird der CELP-Coder (Code
Excited Linear Prediction) zur Codierung von Sprache bei

30

mittleren bis niedrigen Bitraten beschrieben. Dieser
Standard enthält zur Generierung der LPC-Koeffizienten und
der Gain-Werte mehrere Tabellen, die im Gleitkommaformat
vorliegen.

99/807015 ECU



PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

Bescheinigung

DE 99/2633

Die ROBERT BOSCH GMBH in Stuttgart/Deutschland hat eine Patentanmeldung
unter der Bezeichnung

"Verfahren zur Codierung oder Decodierung von Sprach-
signalabtastwerten sowie Coder bzw. Decoder"

am 6. Oktober 1998 beim Deutschen Patent- und Markenamt eingereicht.

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprüng-
lichen Unterlagen dieser Patentanmeldung.

Die Anmeldung hat im Deutschen Patent- und Markenamt vorläufig das Symbol
G 10 L 19/00 der Internationalen Patentklassifikation erhalten.

München, den 13. Oktober 1999

Deutsches Patent- und Markenamt

Der Präsident

Im Auftrag

Weihmayr



Aktenzeichen: 198 45 888.6